Course:ARTIFICIALINTELLIGENCE
TITLE: MARKET BASKET
INSIGHTS
PHASE 2 SUBMISSION

TEAM MEMBERS:
MOHANA PRIYA .M
(au311521106061)
MOUNICKA G.S
 (au 311521106062)
PADMA PRIYA .M.
 (au 311521106069)
YAMINI.H
(au 311521106114)

INTRODUCTION:

In today's data-driven world, extracting valuable insights from vast datasets is a fundamental aspect of making informed decisions in various domains, such as retail, marketing, healthcare, and more. Advanced association analysis techniques and visualization tools play a pivotal role in uncovering hidden patterns, relationships, and trends within data, ultimately leading to enhanced insights and more effective strategies.

Datasource:

• The Online Retail II data set, which includes the sales data of the UK-based online sales store, was used.

• Sales data between 01/12/2009 - 09/12/2011 are included in the data set.

https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis

Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
36365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.550	17850.000	United Kingdom
36365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.390	17850.000	United Kingdom
36365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.750	17850.000	United Kingdom
36365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.390	17850.000	United Kingdom
36365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.390	17850.000	United Kingdom

Association Analysis Techniques: Association analysis, a crucial subset of data mining, helps discover relationships between different variables or items. It is often used for market basket analysis in retail, recommendation systems, and more. The most commonly used algorithm for association analysis is Apriori, which identifies frequent item sets and generates association rules. These rules reveal how certain items are frequently bought or used together. For instance, if a customer purchases coffee, they are likely to buy cream and sugar too, which can guide retailers in bundling products or optimizing store layouts.

Advanced association analysis goes beyond the basic rules to consider more complex relationships. Techniques like lift, confidence, and support thresholds help filter and prioritize the most relevant and meaningful associations. Moreover, association

analysis can be integrated with other data analysis techniques, such as statistical tests and machine learning algorithms, to gain a more comprehensive understanding of the data.

Visualization Tools: Data visualization is a powerful way to communicate complex insights effectively. It transforms raw data into visually appealing charts, graphs, and networks, making it easier for stakeholders to grasp the underlying patterns and trends.

In the context of association analysis, visualization tools are instrumental in presenting the results in an accessible and actionable format. Network analysis tools, such as NetworkX in Python, allow you to create visual representations of association rules as networks. Nodes in the network represent items or variables, and edges represent associations. The thickness of edges or their placement can convey the strength of associations, making it easier to identify influential patterns. These visualizations not only provide a bird's-eye view of complex relationships but also enable quick decision-making.

Visualizations also offer interactive features that allow users to explore data dynamically. They can filter and drill down into specific patterns and even gain insights from the visual representation directly.

The ability to present insights through clear and engaging visualizations ensures that the information is not only understood but also acted upon effectively.

Insights Presentation:
- The combination of advanced association analysis and visualizations empowers data analysts and decision-makers to present insights in a compelling manner.
- Visualization tools facilitate the communication of complex findings and enable quick decision-making.

PROGRAM:

```
import pandas as pd

import networkx as nx

import matplotlib.pyplot as plt

from mlxtend.frequent_patterns import apriori
```

```python
from mlxtend.frequent_patterns import association_rules

# Load your transaction data (e.g., a CSV file with a 'Transaction ID' and 'Item' column)
data = pd.read_csv('transaction_data.csv')

# Perform association analysis
def perform_association_analysis(data, min_support_threshold, min_lift_threshold):
    basket = (data.groupby(['Transaction ID', 'Item'])
            .size()
            .unstack(fill_value=0)
            .reset_index())

    frequent_itemsets = apriori(basket.set_index('Transaction ID'),
min_support=min_support_threshold, use_colnames=True)
    rules = association_rules(frequent_itemsets, metric="lift", min_threshold=min_lift_threshold)

    return rules

# Visualize the association network
def visualize_association_network(rules):
    G = nx.Graph()

    for index, rule in rules.iterrows():
        G.add_edge(rule['antecedents'], rule['consequents'], weight=rule['lift'])

    pos = nx.spring_layout(G)
    labels = {item: item for item in G.nodes()}

    plt.figure(figsize=(12, 12))
    nx.draw(G, pos, labels=labels, width=[data['weight'] for u, v, data in G.edges(data=True)],
node_size=2000, node_color='lightblue', font_size=10)
    plt.title("Association Network")
```

```
    plt.axis('off')

    plt.show()
```

```
# Define your support and lift thresholds

min_support_threshold = 0.1

min_lift_threshold = 1.0
```

```
# Perform association analysis

association_rules = perform_association_analysis(data, min_support_threshold, min_lift_threshold)
```

```
# Visualize the association network

visualize_association_network(association_rules)
```

OUTPUT:

Association Rules:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(A,)	(B,)	0.333333	0.666667	0.333333	1.0	1.5
1	(B,)	(A,)	0.666667	0.333333	0.333333	0.5	1.5

CONCLUSION:

 advanced association analysis techniques and visualization tools are indispensable in data analysis and insights presentation. They empower data analysts and decision-makers to uncover hidden relationships within data, make data-informed decisions, and optimize processes in various industries